12.4 EVALUATION OF THE CONCENTRATION LEVEL UNCERTAINTY ENSEMBLE SYSTEM (CLUES) COMPARED TO ENSEMBLE NUMERICAL WEATHER PREDICTION (NWP) MODEL INPUT USING FIELD TEST DATA

Randolph J. Evans*, W. G. Moore and B. J. Provoncha
ENSCO, Inc., Melbourne, FL

1 INTRODUCTION

The Concentration Level Uncertainty Ensemble System (CLUES) is a tool for estimating uncertainty in modeled atmospheric concentrations derived from transport and dispersion (T & D) models. CLUES comprises several models that attempt to characterize the uncertainty of predicted results from the mesoscale model Regional Atmospheric Modeling System (RAMS) (Pielke et al., 1992; Walko and Tremback 2006) and the dispersion models Short-range Layered Atmospheric Model (SLAM) (Atchison and Kienzle 2002; ENSCO 2008) and CALPUFF. CLUES calculates the uncertainty in these model results using a variant of the Monte Carlo method where winds from RAMS are perturbed with simulated errors. The perturbed winds are then used as input to the selected T & D model. The resulting output can be used to determine the possible errors due to input data uncertainties.

For this evaluation, SLAM dispersion model concentrations were calculated using input meteorological data from CLUES and from an NWP ensemble. The NWP ensemble consisted of a total of eighteen RAMS and Weather Research & Forecasting (WRF) (NCAR 2008) model configurations where model physics options were varied. The modeled concentrations were compared to concentrations collected during a tracer study conducted in the Mojave Desert region during July 2007. Various statistical and graphical comparisons were performed and an assessment of the CLUES and NWP methodologies was conducted. The preliminary results indicate some differences between CLUES and NWP ensembles but neither one did significantly better than the other when compared to measured concentrations. This presentation will present a description of CLUES, the methodologies used in this evaluation and the results of the comparisons.

**Corresponding author address*: Randolph Evans, ENSCO, Inc., 4849 N. Wickham Rd., Melbourne, FL 32940. e-mail: evans.randy@ensco.com

2 CLUES BACKGROUND

The CLUES-RAMS system adds random perturbations to a deterministic RAMS gridded wind field independently at each time period, thus generating individual members for an ensemble of runs for SLAM. Additional perturbations of the original RAMS output wind field are used to generate each additional member of the ensemble, so only a single RAMS run is needed. The method provides an efficient way to generate ensembles of trajectories and concentrations to be used for estimating uncertainty in the model results (ENSCO, Inc. 2001).

CLUES-RAMS does not perturb the winds independently at each of the model grid points, as might be done in a classical Monte Carlo implementation. Instead, an uncertainty model was formulated that allows the simulated RAMS wind errors to be spatially correlated across the grid. The nature of this correlation is similar to that used in the data assimilation process at Goddard Space Flight Center (Dee and Da Silva, 1999 and Gaspari and Cohn, 1999). Qualitatively, this correlation has the property of being large for pairs of grid points that are close to each other, but gradually decreases to zero for pairs of grid points separated by increasingly larger distances. The statistical model used for the simulated wind errors is that of a three-dimensional Gaussian random field having the correlation structure just mentioned. Efficient implementation of this uncertainty model is based on the weighted average method (Oliver, 1995).

A method often used in state-of-the-art weather forecasting that is more realistic, but more computationally demanding, is to produce a true meteorological ensemble system. In this scenario numerical weather prediction (NWP) mesoscale models are run multiple times, using different or perturbed initial and/or boundary conditions or model physics. Therefore, for each member of the ensemble suite, initial errors or differences between members propagate and grow as forecast time increases. This error growth is due to the uncertainty in the initial conditions, model parameterizations and nonlinear nature of the fundamental meteorological model equations as they are solved over time. The spread of the resulting ensemble member solutions can then be used as input to T&D models to quantify uncertainty in concentration fields.

In contrast, the CLUES-RAMS system applies the same level of perturbation independently for each time period, with no possibility of error propagation correlated through time in the wind fields. Also, since the gridded winds are perturbed from a deterministic RAMS run, there is no possibility for higher-order effects from other perturbed meteorological parameters to influence the resultant wind field through the complex model equations as in true mesoscale model ensemble systems. The tradeoff is in RAMS execution time, which must be multiplied by the number of members in the true ensemble in comparison to CLUES, which uses only a deterministic RAMS run. The use of multiple RAMS runs in CLUES would be prohibitively expensive in terms of execution time for most operational applications. It should be noted that even though errors do not propagate forward in time within the perturbed wind fields, CLUES does allow error propagation within SLAM, since those models are run separately for each member of the ensemble.

There are two additional sources of error that are not considered by CLUES, since they cannot be simulated when using only a single RAMS run. One is the effect of errors in the initial and boundary conditions, which will produce changes in the wind field as RAMS makes use of data from outside the initial grid. This effect can be simulated by perturbations in the boundary conditions, which must be different for each RAMS simulation. A second source of error is the physical model used for generation of wind fields. Its effect can be estimated by perturbing various parameters and options in the model. However, most state-of-the-art meteorological ensemble systems focus on the “free” forecast problem where the NWP model is initialized using a set of observations to form a “best estimate” of the true atmospheric state and then run out to a given time. In the current application, RAMS is continually nudged to both gridded analyses and observations (surface and upper air) to form the “best” re-creation of the actual meteorological record. Therefore, the errors that grow considerably large in “free forecast” mode are somewhat constrained by the observations and analyses that are brought into the model at given times. Even though the gridded analyses and observations constrain the RAMS simulation compared to reality, error propagation still occurs through the uncertainties related to the model physics and dynamics, lateral boundary conditions, data assimilation scheme, and the observations themselves.

The goal of this task was to compare the results from CLUES-RAMS to a full mesoscale model ensemble system.

3 METHODOLOGY

The methodology for the CLUES validation was based on the data collected during a tracer field test which is described in detail in Tracer Environmental Sciences & Technology, Inc. (2007).

3.1 Tracer study

An atmospheric tracer study was conducted in July 2007 in the Mojave Desert region. Three different perfluorocarbon (PFT) tracers were released from three different locations for release durations of four hours. Twenty-five samplers were arranged in two downwind arcs and collected samples in 30-minute and 60-minute increments. A map showing the source and sampler configuration is shown in Figure 1.

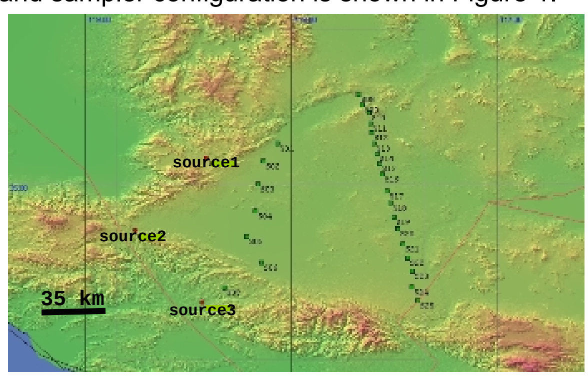


Figure 1. Map of the SLAM domain used in the tracer test in the Mojave Desert region. The three source locations are shown by the red circles and the various sampler locations are denoted by the green squares.

3.2 CLUES Configuration

A baseline RAMS run was made that covered the period of the tracer Test 2 and another RAMS run was made that covered the period of the tracer Tests 3 and 4. The physics and grid options selected for the RAMS baseline were based on previous runs made for this area. CLUES-RAMS was run with an ensemble of 50

CLUES members to provide sufficient convergence of the results.

Once the CLUES-RAMS runs were completed, the CLUES-SLAM runs were made. These CLUES-SLAM runs used the 50 members of the CLUES-RAMS runs as input meteorological data. The runtime configurations for the CLUES runs for the validation study are shown in Table 1.

3.3 NWP Configuration

In order to design a realistic ensemble system multiple uncertainties in the model must be examined. This includes the model initial and boundary conditions, data assimilation schemes, model physics, and land surface interactions. The ensemble methodology used here follows that of the Mesoscale Model 5 (MM5) system described in Liu et al. (2007). They found that there was an advantage of multi-model ensembles based on varying the dynamics formulations such as in using two different mesoscale models. They found that the physics diversity and other perturbation approaches appear to be less dispersive, though important also.

Table 1. Runtime configurations for CLUES runs for validation study.

Run	RAMS start	RAMS duration (hrs)	CLUES-RAMS start	CLUES-RAMS duration (hrs)	CLUES-SLAM start	CLUES-SLAM duration (hrs)
Test 2	14 July 0000 UTC	84	16 July 1900 UTC	13	16 July 1930 UTC	12
Test 3	17 July 1200 UTC	96	18 July 1600 UTC	16	18 July 1630 UTC	15
Test 4	17 July 1200 UTC	96	20 July 2000 UTC	14	20 July 2000 UTC	14

Currently, RAMS has a limited number of physics options and data assimilation choices. Given these limitations a multi-model ensemble system was developed using different configurations of RAMS and the Weather and Research Forecast (WRF) model to generate the uncertainty of using a single RAMS run. This is often done in the modeling community and is quite common in the literature. Using more than one mesoscale model allows for many more perturbation methods (e.g. physics and data assimilation) to be applied to the ensemble enabling a far more robust system.

To address the uncertainties associated with using a single mesoscale model run, a number of perturbation methods were applied in the multiple-model ensemble approach. A list of the perturbation methods used in the RAMS/WRF ensemble system is outlined in Tables 2 and 3. It was important to keep the number of RAMS and WRF members in the ensemble about equal to avoid one model from causing bias in the outcome. Additionally, while constructing the multiple-model ensemble system it was important not to choose schemes that are inferior to others therefore biasing the results (e.g. comparing dry microphysics to the full mixed-phased schemes). To satisfy both these requirements and to complete the task within the time allotted, an 18 member ensemble system was developed; nine members from each model were developed. A key assumption in an ensemble is that each member outcome is an equally likely scenario as the next. For the analyses in this study the baseline RAMS run was considered member one in the NWP ensemble system.

Table 2. List of physics variations used for RAMS ensemble runs.

RAMS
Global Forecast System (GFS) Initial Conditions (IC) and Boundary Conditions (BC) with standard options
GFS IC and BC with data assimilation weighting tweaked upward
GFS IC and BC with data assimilation weighting tweaked downward
GFS IC and BC with data assimilation radius of influence doubled
GFS IC and BC with Kuo convection scheme
GFS IC and BC with LEVEL=2 microphysics scheme 1
GFS IC and BC without 1-km inner nest (i.e. 5 km)
GFS IC and BC with Mahrer/Pielke shortwave/ longwave- radiation scheme 2
North American Reanalysis IC and BC with standard options

Table 3. List of physics variations used for WRF ensemble runs.

WRF
GFS Initial Conditions IC an BC with standard options
GFS IC and BC with data assimilation weighting tweaked upward
GFS IC and BC with data assimilation weighting tweaked downward
GFS IC and BC with data assimilation radius of influence doubled
GFS IC and BC with Betts-Miller convection scheme
GFS IC and BC with WSM microphysics scheme
GFS IC and BC with Mellor Yamada Planetary boundary layer scheme
GFS IC and BC with CAM3 shortwave/longwave radiation scheme
North American Reanalysis IC and BC with standard options

3.4 SLAM Configuration

The configuration of the SLAM runs were matched with the actual tracer Tests 1, 2, and 3 so that the release times, release rates, and sampling times agreed with actual release and sampling data. Source and sampler locations were the same and were shown in Figure 1.

Concentration data were computed as hourly averages to match the collection period of the samplers. The samplers collected hourly samples on the outer ring of samplers (Samplers 8 through 25) and half-hourly samples on the inner ring (Samplers 1 through 7). To keep sampling periods uniform and to match SLAM modeled concentrations with actual sampled concentrations, the half-hour samples were merged into hourly samples.

3.5 Statistical Tests

Because the 25 samplers collected samples for a relatively short time period (8-12 hours), it was decided to conduct some of the comparisons between modeled and collected concentrations using the sum of the concentrations for the full length that the sampler was collecting for each of the three tests. The time-integrated concentrations with example units of g-hr/m^3 were computed by summing the concentrations over the period of sampling to compute a total concentration in units of g/m^3. The summed concentrations were adjusted to account for the half-hour samples that were collected in the inner ring of samplers.

There have been many papers that address the quantitative comparison of model predictions to observational data. Different statistics have been proposed that highlight the central tendencies and variability of these model/data comparisons. Some of these statistics are calculated using scaled data because of the fact that concentration or sum concentration data very often differ by orders of magnitude. We decided to apply four of these established and accepted statistics, along with one additional statistic, to the comparisons of the CLUES and NWP data.

Although the main interest for this report is the comparison of the CLUES data to the NWP (9 RAMS runs and 9 WRF runs) data, graphs were created that compared the RAMS and WRF ensembles along with the data from the 50 CLUES member runs. Doing this gives some idea of the variability within each of the different NWP ensembles and whether there were any major differences between the two NWP ensembles with respect to the five statistics studied.

There is a certain intuitive appeal to statistically comparing data from competing models where these statistics are based on how these models relate to real observational data (truth). These observational data give a reference point from which the comparisons between the CLUES and NWP data can be made. Without a reference point, differences between the two models would be difficult to interpret; even if there were differences between the two models, there would be no way of saying which model was more accurate.

When making these types of comparisons, observation and model data can be paired in space, time or both space and time. A decision was made to pair the observed and modeled data in space. For a given sampler, sum concentrations were created by integrating the concentration data over time. One unintended consequence of temporally combining the data is the reduction of the effective sample size. This becomes important when calculating the variability of the different statistics because less data implies more variability.

There were many observed concentrations that were missing, and the vast majority of these missing data were non-detects (concentrations below the detection limit of the sampler). Adding the data over time helped this condition to a degree, but there were still many data points left with no numerical value. These data were deleted from the observation dataset before being paired with the modeled sum concentrations. A similar problem existed with the modeled data; there were many zeros. Granted, the majority of these values were probably true (modeled) zeros, but some of these would also be values that fell below the modeled detection limit. With the exception of one of the calculated statistics, Measure of Effectiveness (MOE), modeled data having a sum concentration value of zero were also deleted prior to pairing with the observed sum concentrations. Thus all of the paired data used for calculating the various statistics consisted of observed and modeled sum concentrations with non-zero values.

The final paired sum concentration dataset consisted of data from all three tracer tests and all three sources combined for all 25 samplers. One could argue for separating the data according to test and source, but doing this would have made the statistical comparisons of CLUES and the NWP very questionable because the dataset sample sizes would have been too small for reliably calculating the variability of the statistics.

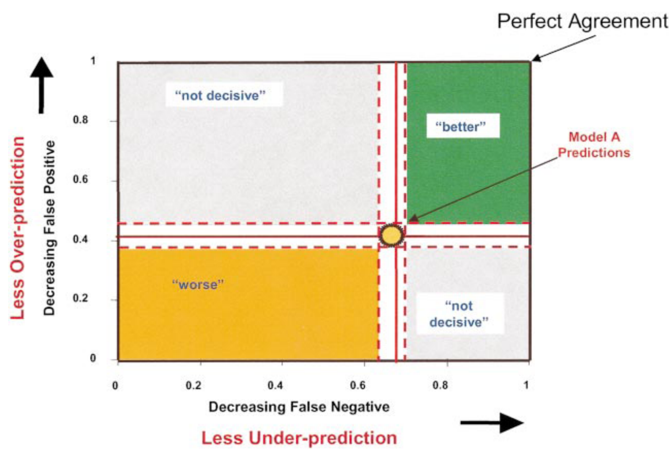


Figure 2. Explanation of Measure of Effectiveness (MOE) graph from Warner et al (2004).

Confidence intervals for each of the statistics were calculated using the method of bootstrapping (Efron and Tibshirani, 1993). This is a popular technique for calculating standard errors and confidence intervals for statistics that have no handy analytical formula for calculating them. The simplest form of the bootstrap is performed by sampling the dataset with replacement and calculating the statistic of interest. This re-sampling is done many times (1000 or more), and a distribution of statistics is the final product. There are different ways of calculating confidence intervals from this distribution of statistics, and the method that was used for these data is called the BC_a method, which stands for bias-corrected and accelerated. With sufficient sample size, this method gives accurate confidence intervals with respect to coverage and probability.

The first two statistics that are listed below come from Chang and Hanna (2004). FAC2 stands for the fraction of the modeled sum concentration that were within a factor of two of the observed sum concentrations. MG is called the geometric mean bias, and it is simply the ratio of the geometric mean of the observed data to the geometric mean of the modeled data; geometric means are useful for describing log-normally distributed or skewed data. The third statistic listed, FAC10, is a variant of FAC2, and it is simply the fraction of the modeled sum concentrations that were within a factor of ten of the observed sum concentrations:

FAC2 = fraction of data that satisfy $0.5 \leq \frac{C_p}{C_o} \leq 2.0$, (1)

MG $= \exp(\overline{\ln C_o} - \overline{\ln C_p})$, (2)

FAC10 = fraction of data that satisfy $0.1 \leq \frac{C_p}{C_o} \leq 10.0$, (3)

where:

C_p: modeled sum concentrations,

C_o: observed sum concentrations,

$(\overline{\ln C})$: average of the log-transformed values over the dataset.

If the model perfectly predicted the observed data, all three statistics above would be equal to one. The FAC and MG statistics attempt to show how close the observed and modeled sum concentrations match up in space, and MG can give an idea of directional bias. There was another statistic studied that had the same goal:

MEDRATIO $= Median(\frac{C_o}{C_p})$, (4)

This is the median of the ratios of observed to modeled sum concentrations. This statistic can also determine whether the modeled sum concentrations tend to over-predict or under-predict the observations. The median is used instead of the mean because of the wide spread of the ratios. A perfect model would also have MEDRATIO = 1.

One final statistic that was used to compare the CLUES sum concentrations to the NWP sum concentrations is called a Measure of

Effectiveness (MOE) by Warner et al (2004). This is a two-dimensional statistic that is shown on a Cartesian graph, and the two components are the false-positive fraction (over-prediction) and the false-negative fraction (under-prediction). Figure 2 was taken from Warner et al (2004), and it illustrates some aspects of the MOE graph. Both axes are labeled, and their scale is from zero to one. The gold circle in the figure represents an estimate of the MOE for some set of data (Model A). If another estimate of the MOE was made and compared to Model A's prediction, the four shaded areas show how the new estimate would compare in relation to Model A: green area would be better, orange area would be worse, two gray areas leave room for subjective interpretations. Of course, the boundary lines themselves are subjective, but they serve as a good general rule.

The MOE statistic would be shown as one point on the graph, but the bootstrapping procedure produces a cloud of points showing the likely variability of the statistic. The clouds from the CLUES and NWP statistics will be statistically compared by looking for cloud separation between the models.

The MOE is the only statistic used in this study that was calculated for each row of samplers separately (samplers 1 through 7 and samplers 8 through 25). This had to be done because of the way that the statistic is calculated. Data for all three tests and all three sources were still combined to calculate the MOE for each set of samplers.

These five statistics are certainly not an exhaustive list for comparing the CLUES and NWP data, but they should be satisfactory at testing whether the central tendency of the CLUES and NWP data are different.

4 RESULTS

4.1 Trajectory comparisons

For the trajectories that were created for this study, maps were produced to compare the CLUES-SLAM trajectories versus the NWP-SLAM trajectories. Additionally, the NWP-SLAM trajectories were color-coded to indicate the difference between the RAMS-SLAM and WRF-SLAM trajectories since the trajectory results showed that there was a marked difference between RAMS and WRF trajectories in this study.

Figure 3 shows examples of trajectories from some of the runs. Some of the key results were:

- RAMS trajectories from source 1 consistently travel left of the samplers and left of WRF trajectories for all three tests, possibly due to the models not handling terrain interaction with the flat desert to the south and higher, rougher terrain to the north.
- Comparing trajectories from source 2 indicated that CLUES and NWP trajectories overlapped
- A few of the RAMS and CLUES trajectories occasionally went to the west from sources 2 and 3 while the majority of the trajectories for those runs went to the east. The westward movement of the few trajectories was opposite the direction of most of the trajectories.
- Comparing the dispersiveness of the trajectories, CLUES trajectories generally covered an area as wide or even wider than the NWP trajectories. The difference was primarily due to CLUES runs contained 50 members while NWP runs contained only 18 members. When the NWP trajectories covered an area not covered by the CLUES trajectories it was when the WRF trajectories differed distinctly from the RAMS trajectories.

4.2 Statistical results

The comparison of the NWP and CLUES RAMS ensembles is performed by looking at all three MOE graphs within *Figure* 4. The considerable variability present in all three graphs does not allow any differences to be detected statistically. The variability seen in these MOE graphs was due to a combination of having relatively few samplers to calculate the MOE statistic and having only nine records (3 tests and 3 sources) available for the bootstrapping algorithm.

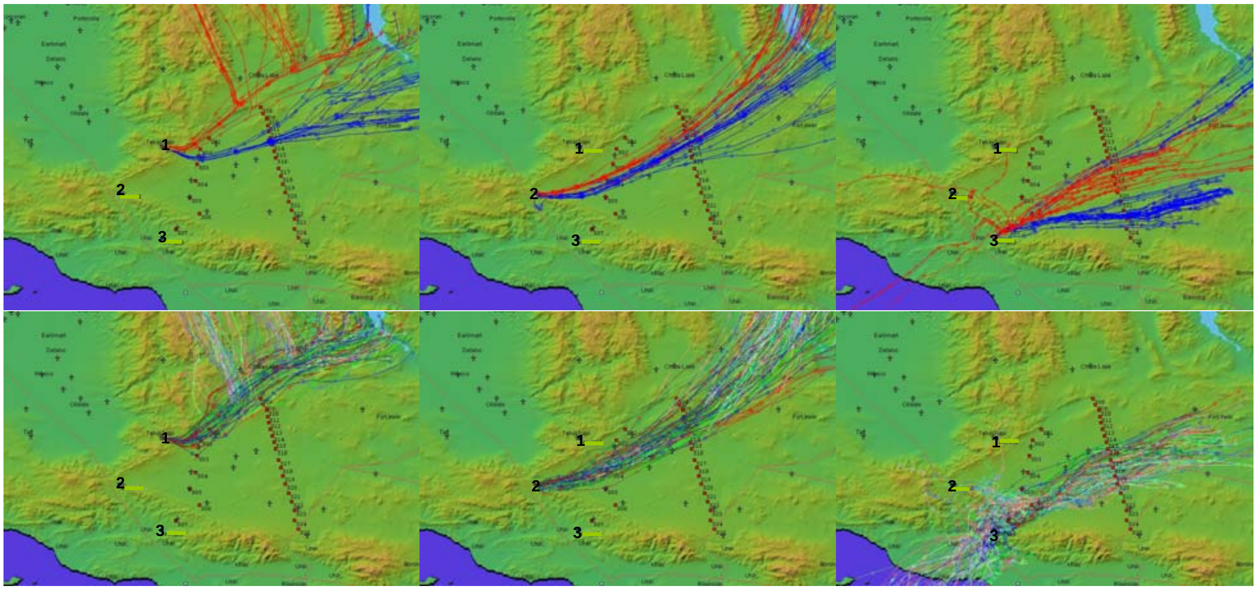
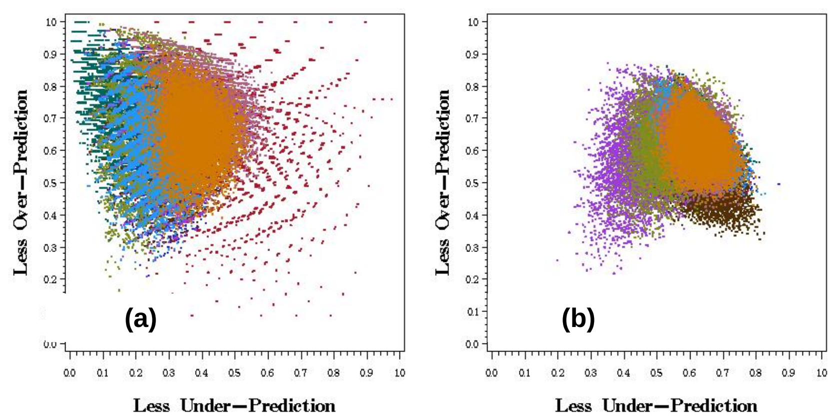


Figure 3. SLAM Trajectories from NWP (top) and CLUES (bottom) for Test 2, Source 1, 2330 UTC, Test 3, Source 2, 1830 UTC, Test 4, Source 3, 0000 UTC. In the NWP figure, red trajectories are from RAMS and blue trajectories are from WRF.



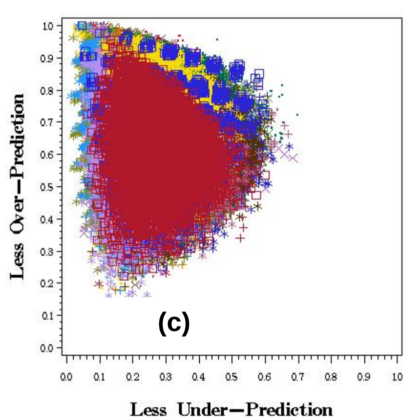
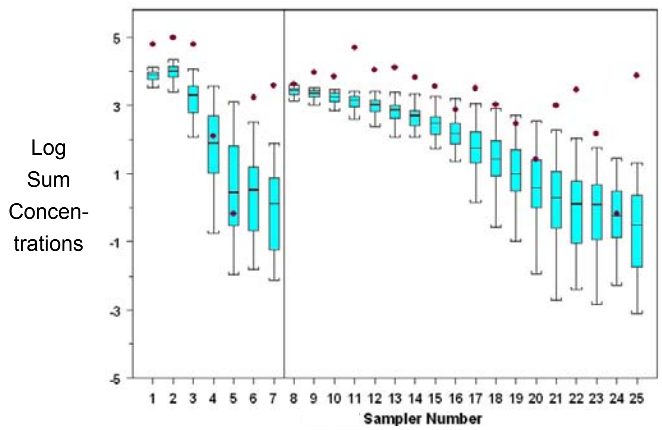


Figure.4. MOE clouds for: (a) NWP RAMS, (b) NWP WRF PIMIX, (c) CLUES RAMS. The different colors within each graph are for the individual members (9 for the two NWP graphs and 50 for the CLUES RAMS). Samplers 8 - 25.

4.3 *Box and Whisker Plots*

As a way to compare the CLUES-predicted data and the NWP-predicted data and to visualize the range of concentrations produced by the different ensemble members, box and whisker plots were produced for each test, each source and each sampler. These plots are a good way to determine if the ensemble members were able to encompass the actual concentrations when looking at the sum concentration at each sampler. The box and whiskers show the median concentration along with the 25^{th}-75^{th} percentile, and the effective range. An example box and whisker plot is presented in Figure 5. The box and whisker plots show logarithmic sum concentrations.



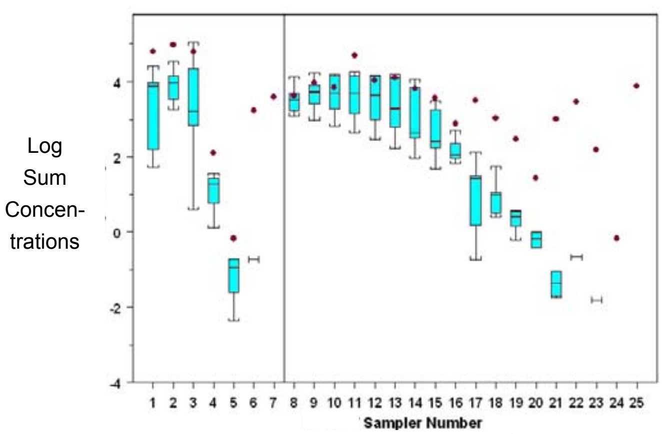


Figure 5. Box and whisker plots (5, 25, 50, 75, 95 percentiles) for 50 CLUES members and 18 NWP members at each of the 25 samplers compared to the observed concentration sum (●) for Test 2 Source 1. Sum concentration scale is logarithmic base 10.

Therefore, when comparing the differences between two different concentrations of higher percentiles located in the top of the graphs there is a much greater actual difference than comparing the difference between two lower percentiles located in the lower part of the graphs.

5 CONCLUSIONS

The statistical results showed that there were no significant differences between the NWP and CLUES results with respect to the five statistics studied. There was considerable variability with the data that precluded detection of any differences, if in fact there were any true differences between the NWP and CLUES RAMS. Relatively low samples sizes contributed to the width and size of the confidence intervals and confidence clouds.

Box and whisker plots were used to show the variations of the NWP and CLUES ensembles compared to observed concentrations. When comparing predicted to observed concentrations that were summed over the tracer experiment duration we found that the CLUES ensembles were not significantly different from the NWP ensembles. There were several exceptions and these were:

- **NWP predicted concentrations in general produced more concentration variation than CLUES.** The box and whisker plots for the NWP runs were generated from the sum of the concentrations from 18 different ensemble members while the CLUES plots were from 50 ensemble members. The wider variation of the NWP plots was most likely due to the NWP runs were based on ensembles of two different models—WRF and RAMS, whereas the CLUES ensembles were based on wind perturbations of just one model—RAMS.

- **For the most northern of the three sources, source #1, for all three tests, CLUES under-predicted concentrations compared to observations**. The reason for the underprediction was because many of the CLUES trajectories missed the sampler rings and traveled to the north of the samplers. NWP runs did better than CLUES at matching the observed concentrations because the NWP runs included the WRF data as input and the WRF-based trajectories traveled toward the east from source #1 into the samplers. . These differences between CLUES and NWP were due primarily to RAMS poorly modeling trajectory directions from source #1 and causing CLUES-modeled trajectories, which are based on RAMS trajectories as its baseline, to inaccurately miss the samplers.

- **For the middle of the three sources, source #2 for all three tests, both NWP and CLUES ensembles produced concentration patterns that matched observations as indicated by the box and whisker plots**. The plots showed that the concentration range between the 18 NWP runs and 50 CLUES runs were distributed similarly between lowest and highest concentration and that for most of the samplers the observed concentration fell within that distribution.

- **For the most southern of the three sources, source #3, the NWP runs showed slightly larger concentration variation than the CLUES runs.** This difference is probably attributable to the fact that the NWP runs were derived from two different models and the CLUES runs were derived from one model. Compared to sum of observed concentrations for the source #3 runs, neither the NWP or CLUES did significantly better that the other.

When comparing spread of trajectories between CLUES and NWP runs, the trajectory plots indicated that spatially they were very similar. The differences that were noted were due to the NWP WRF runs which varied from the NWP RAMS and the CLUES trajectories. This result was not surprising because the CLUES runs use RAMS as its baseline run.

Because the three tracer tests were conducted mostly during the daytime (all releases occurred during the daytime and sampling continued into nighttime in two of the tests), we were not able to compare diurnal differences between CLUES and NWP in this evaluation. In earlier study, we found that CLUES and NWP trajectories were similar during the nighttime but tended to differ somewhat during the daytime. We also found both CLUES and NWP trajectories at the sources in the rugged terrain were more strongly influenced by the terrain in the nighttime and early morning stable conditions than in daytime unstable conditions.

The WRF and RAMS ensembles were generated by varying the input model physics of each model based on other researchers' studies. Eighteen NWP model runs were generated (nine WRF and nine RAMS). To summarize the conclusions and to answer the question of whether CLUES ensembles or NWP ensembles do a better job of capturing the uncertainty for concentration predictions when using observed concentrations as the basis for the comparison we found the following:

- CPU and wall clock time to generate CLUES ensembles is considerably less than NWP ensembles (hours vs. days) and disk space required is less as well. For these runs CLUES data used approximately 50 gigabytes of disk space for a 50-member ensemble versus approximately 650 gigabytes for the 18-member NWP ensemble.

- NWP produces wider variation in predicted concentrations than CLUES and its variability incorporates observed concentrations slightly better than CLUES. The wider variation in NWP results is attributed primarily to using two different models.

- CLUES ensemble trajectories, in general, favorably compare with the NWP ensemble trajectories.

- Occasionally but not always, at different times and/or different sources, RAMS ensembles differed from WRF ensembles in both trajectories and concentrations predictions. However, the variation within the nine WRF model results and within the nine RAMS model results were not that different.

- There was a tendency for the all of the models used in this study to under-predict the observed data.

6 REFERENCES

Atchison, M. K., and M. A. Kienzle, 2002: Utilizing mesoscale model output within the SLAM -P model framework. *12th Joint Conf. on the Applications of Air Pollution Meteorology with the AWMA*, Amer. Meteor. Soc. 2pp.

Chang, J. C., and S. R. Hanna, 2004: Air quality model performance evaluation. *Meteor. Atmos. Phys*., **87**, 167-196.

Dee, D.P. and A.M. da Silva, 1999: Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology. Mon. Wea. Rev., 127,1822-1834.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*, No 57, Chapman and Hall, 436 pp.

ENSCO, Inc., 2001: Ensemble Uncertainty Analysis (EUA) User's Guide, Prototype Version 0.1. by ENSCO Inc., 4849 Wickham Rd., Melbourne, FL 32940

ENSCO, Inc., 2008: Technical Documentation for the Short Range Layered Atmospheric Model SLAM Version 4.20 Geosystem Solutions Division, 4849 Wickham Rd., Melbourne, FL 32940 June 2008, 144 pp.

Gaspari, G. and S. Cohn, 1999: Construction of Correlation Functions in Two and Three Dimensions. Quart. J. Roy. Meteor. Soc., 125, 723-757.

Liu, Y., M. Xu, G. Roux, M. Ge, R. Sheu, J. Hacker, T. Warner and S. Swerdlin, 2007: A FDDA based operational mesoscale ensemble analysis and prediction system (E-RTFDDA) and a comparison of MM5 and WRF. 2007 WRF Users Workshop. Boulder, CO. 8 pp.

NCAR, 2008: User's Guide for Advanced Research WRF (ARW) Modeling System Version 2.2, National Center for Atmospheric Research, 226 pp.

Oliver, D.S., 1995: Moving Averages for Gaussian Simulation in Two and Three Dimensions. Math. Geol., 27, 939-960.

Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee, and J.H. Copeland, 1992: A Comprehensive Meteorological Modeling System – RAMS. Meteor. Atmos. Phys., 49, 69-91.

Tracer Environmental Sciences & Technology, Inc., 2007: Field Study Support Study Project Report, San Marcos, CA.

Walko R. L. and C. J. Tremback, 2006: RAMS ,Regional Atmospheric Modeling System, Version 6.0 User's Guide – Introduction, 8 pp. Available online at http://www.atmet.com/html/docs/documentation.shtml.

Warner, S., Platt, N., and J. F. Heagy, 2004: User-Oriented Two-Dimensional Measure of Effectiveness for the Evaluation of Transport and Dispersion Models. *J. Appl. Meteor*., **43**, 58-73.